

Surveying Leading Open-source LLMs and Their Use Cases



Laurentiu Raducu

Founder of bitheap

@bitheap_tech | www.bitheap.ch

Which model should you choose?



Summary

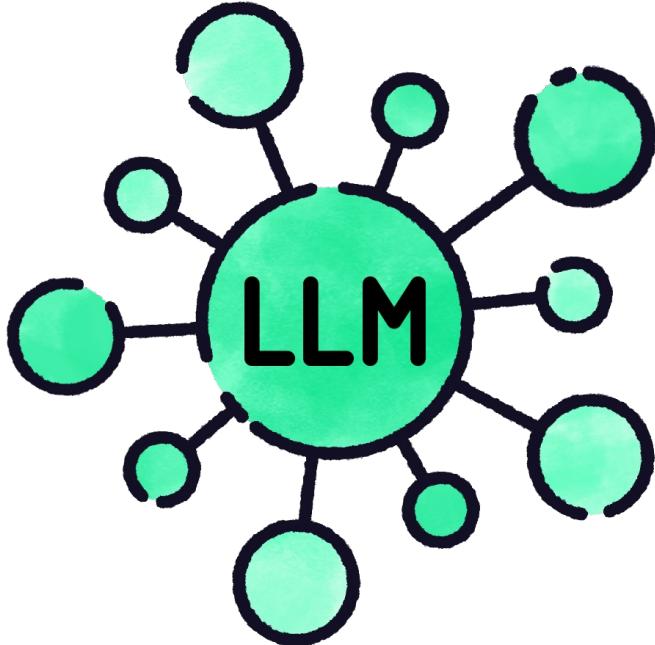
Overview of leading open-source LLMs

Learn key architectural innovations

Analyze performance on standard benchmarks

Learn to match each model's strengths to specific business cases

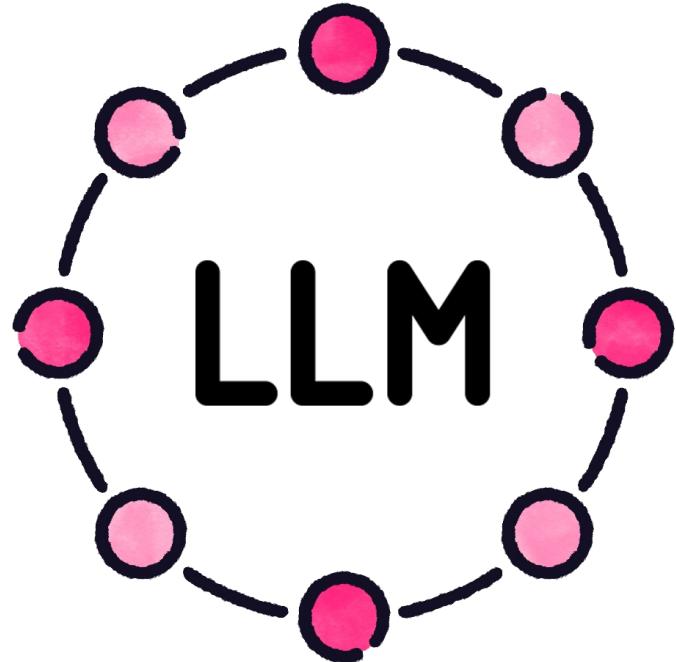




LLaMA 4 characteristics:

- Not fully “open source”
- High general-purpose performance





BLOOM characteristics:

- Collaborative project involving over 1000 researchers
- Multilingual
- Responsible AI license

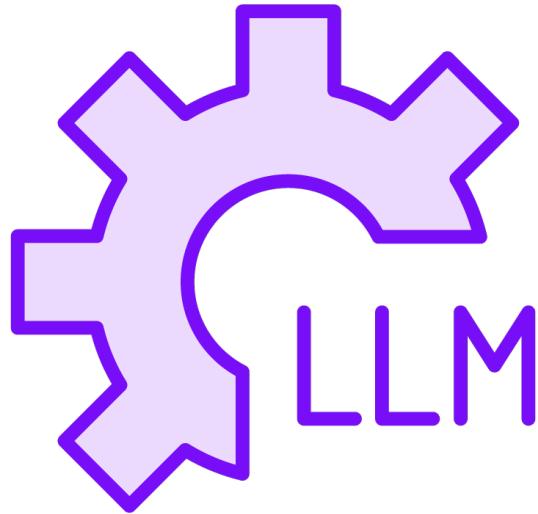




DeepSeek characteristics:

- Base models outperformed much larger models
- Excels at reasoning, coding and math

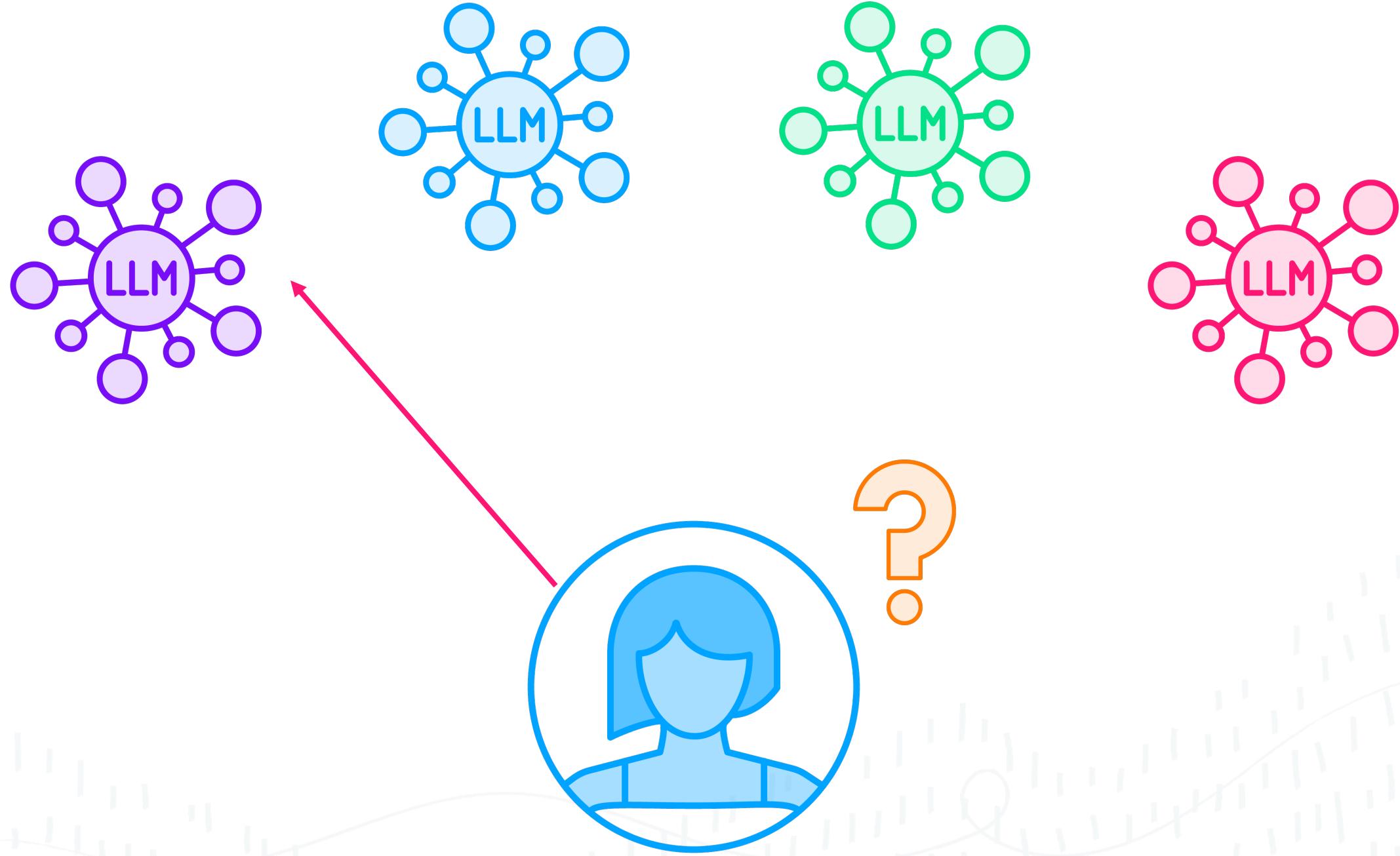




Mistral characteristics:

- Competes on efficiency
- Uses Apache license

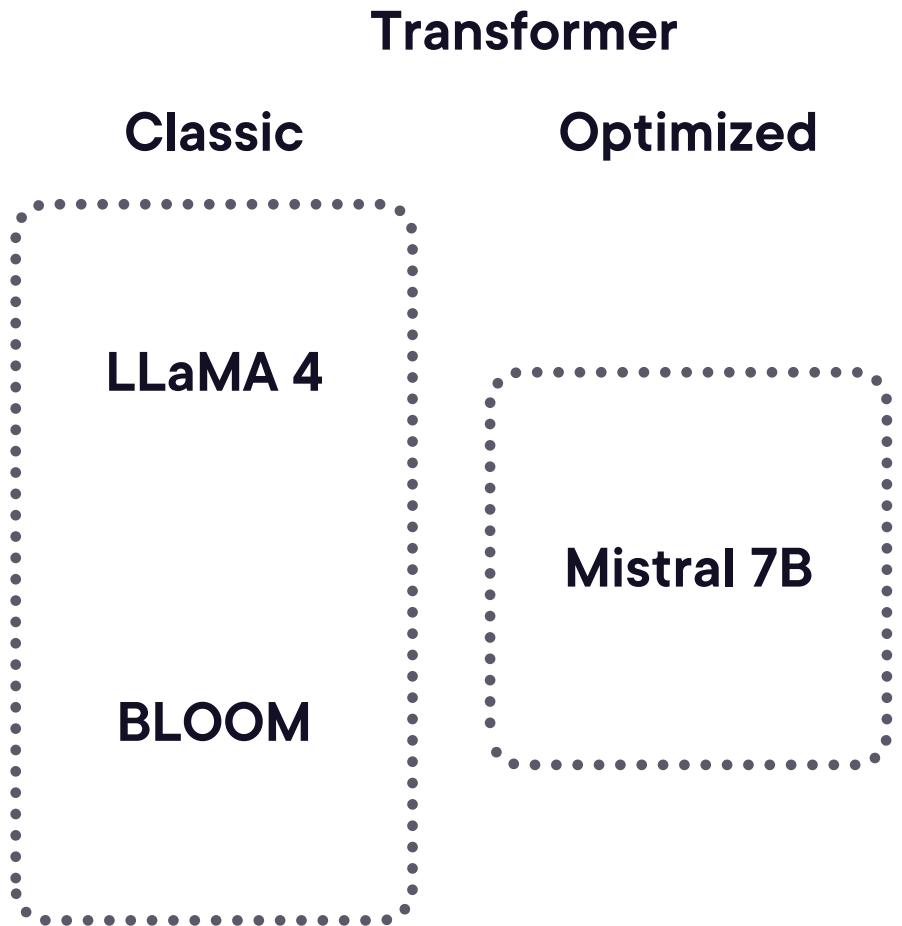


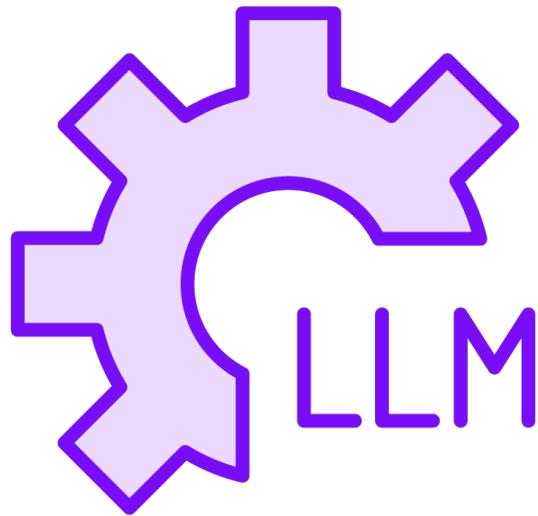


Comparative Analysis: Architecture, Scale and Benchmarks



LLM Architectures



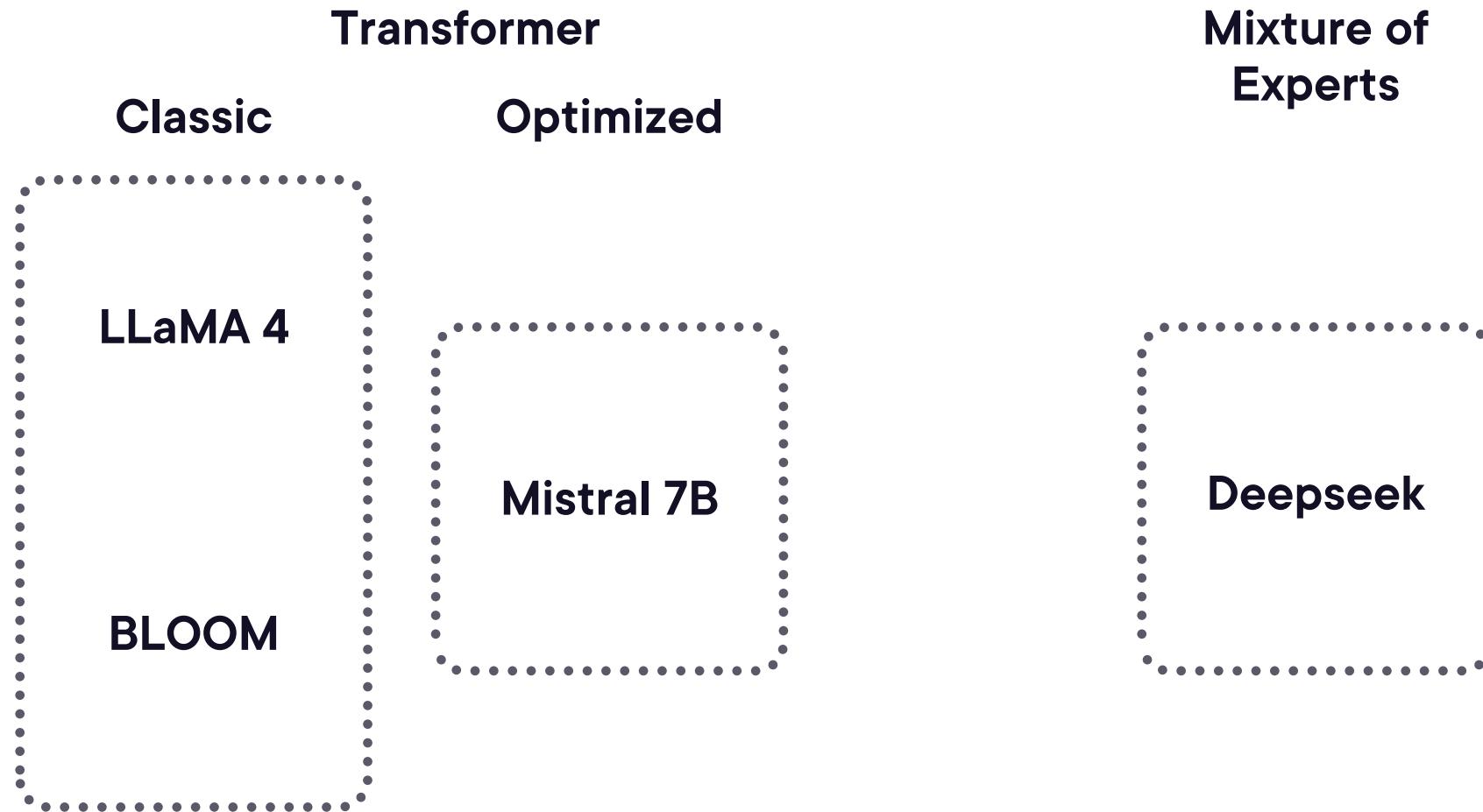


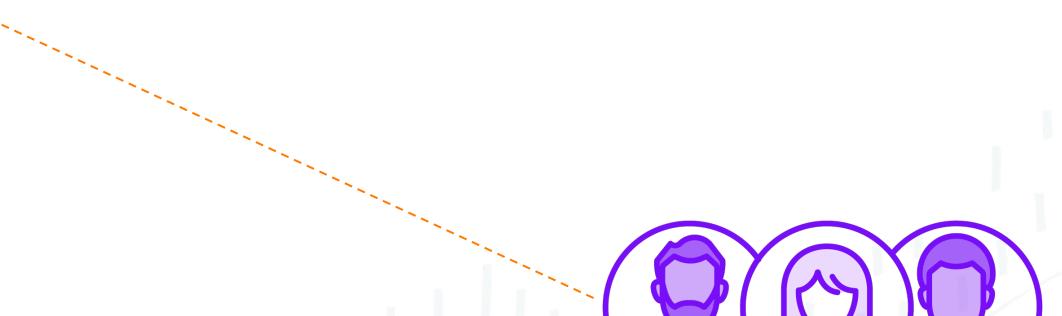
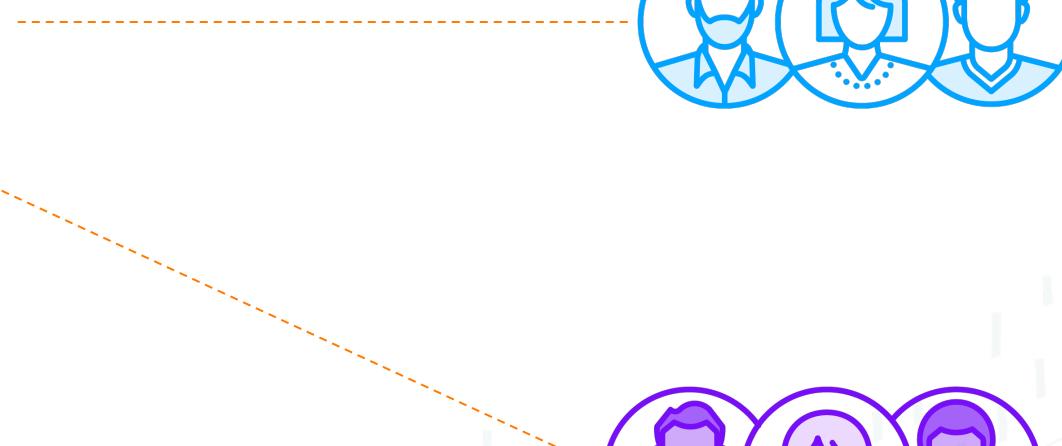
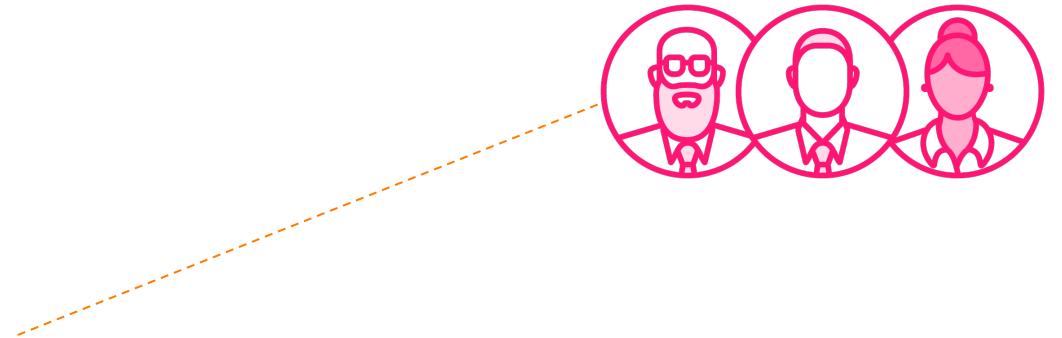
Mistral optimization techniques:

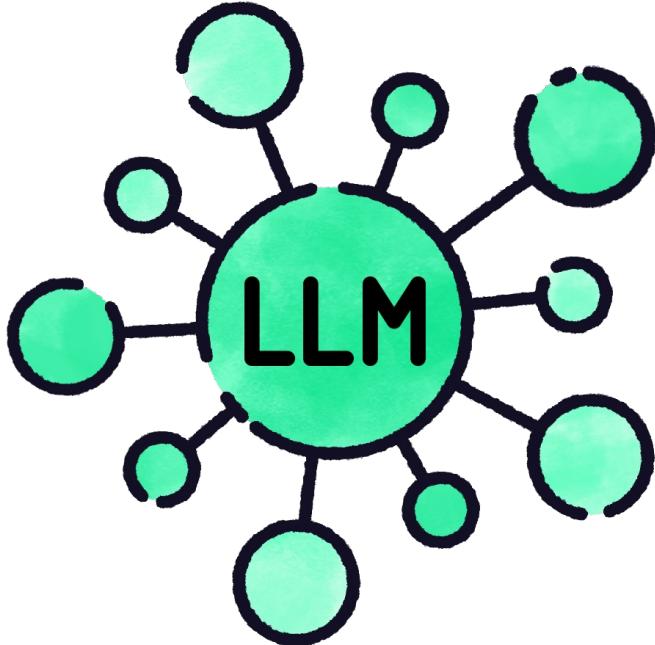
- Sliding window attention mechanism
- Grouped-query attention



LLM Architectures







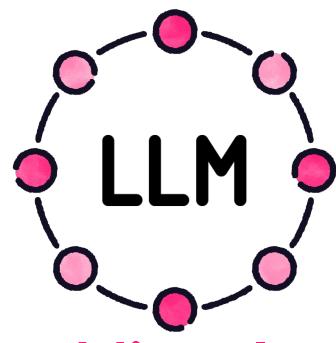
LLM scale

- Parameter count shows the model's scale
- More parameters = greater capacity for knowledge and nuance

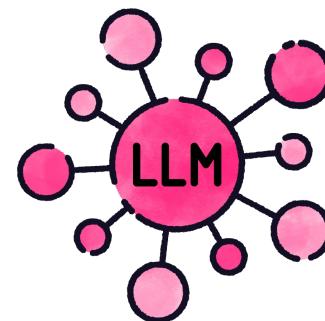


Scale of Open-source Language Models

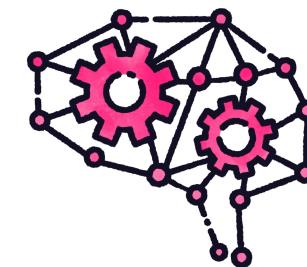
A summary of the parameter count for 3 different open-source models



7 billion parameters



67 billion parameters



671 billion parameters



**How do these architectural
and scale differences
translate into real-world
capability?**



Useful Benchmarks

Massive Multitask Language Understanding, [General Knowledge](#)

GSM8K, [Grade-school Math Reasoning](#)

HumanEval, [Code Generation Benchmark](#)

MT-Bench, [Conversational Ability](#)



Model Ranking

DeepSeek-67B



LLaMA 4 70B



Model Ranking

Mistral 7B



LLaMA 2 13B





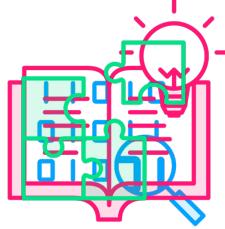
Specialized Models and Optimal Use Cases



Optimal Use Cases

Advanced Reasoning

DeepSeek-R1



DeepSeek-R0

Software Development

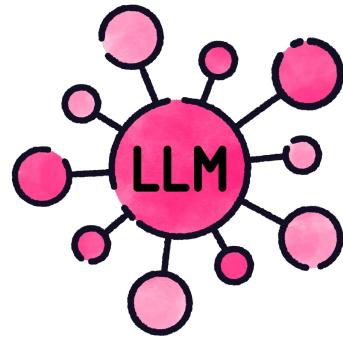
DeepSeek-Coder-V2

- Supports 338 programming languages
- 128,000 tokens

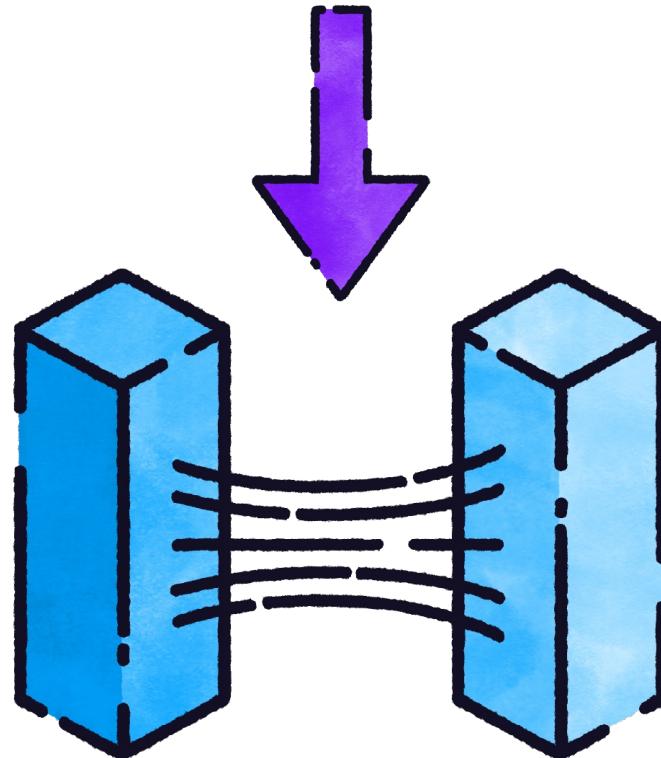


Practical Deployment





Mistral 7B



Optimal Use Cases

Advanced Reasoning

DeepSeek-R1

DeepSeek-R0

Software Development

DeepSeek-Coder-V2

- Supports 338 programming languages
- 128,000 tokens

Practical Deployment

Mistral 7B

- Best for efficiency

- Apache 2.0 license



Which Model to Choose?

DeepSeek-R1 for
multi-step reasoning

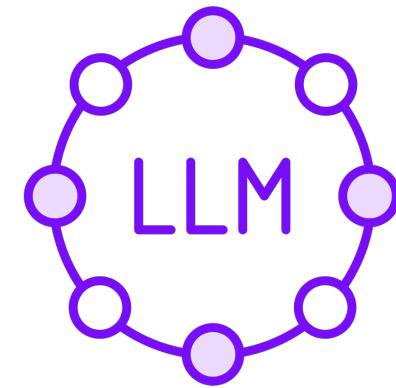
DeepSeek-Coder-V2
for software
development

Mistral 7B for
running the model
on-premises with
limited hardware



Model Strengths, Weaknesses, and When to Use Each





LLaMA 4

Versatile and strong general-purpose performance

Supported by a massive community

Does not have a permissive license

Not a top performer for coding and reasoning



BLOOM

Great multilingual capability

Community-driven project

Lags behind newer models in coding and reasoning

RAIL license has restrictions



DeepSeek

**Excels in reasoning, math, coding
and Chinese language
comprehension**

**Outperforms much larger
competitors**

Complex to deploy

Has a custom commercial license



Mistral 7B

Efficient

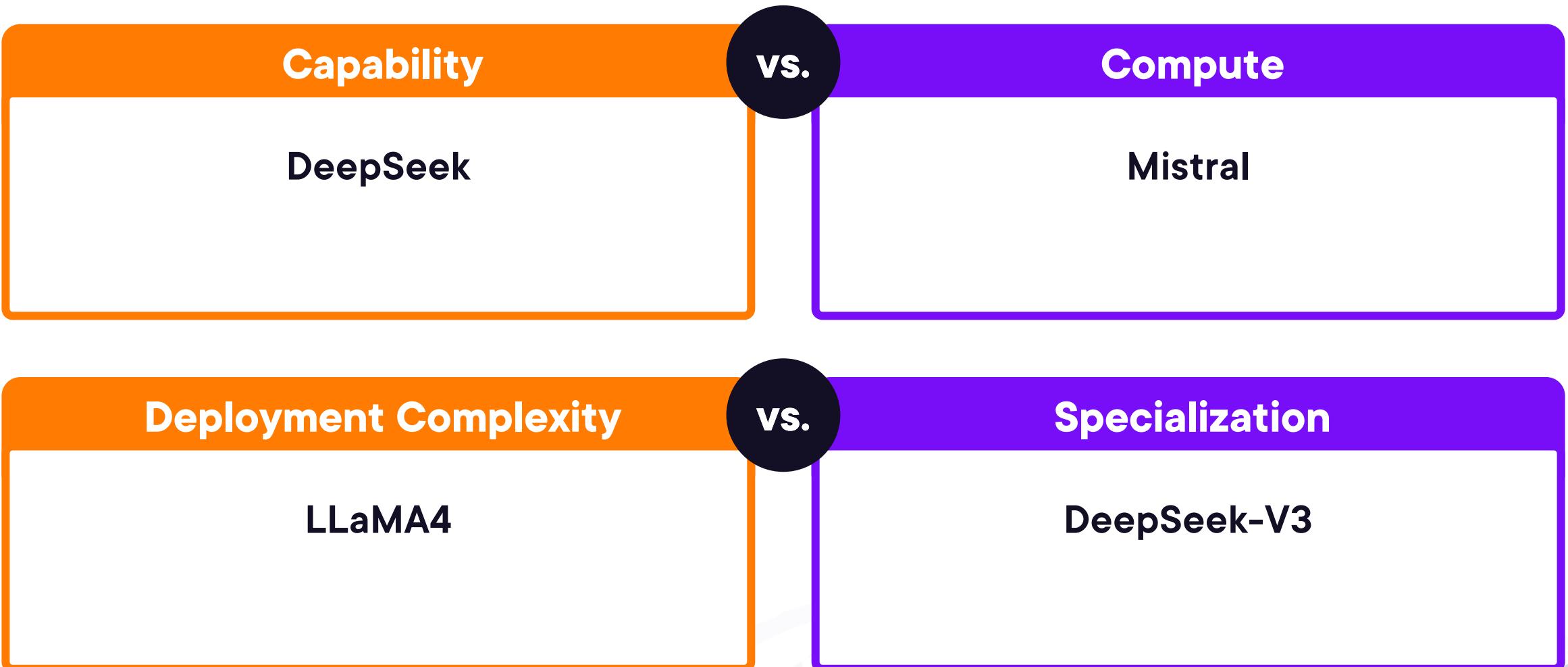
Apache 2.0 license

Great for self-hosting

Lacks deep knowledge than 70B+ models



Balancing Trade-offs



Types of Licenses

Openness

vs.

Licensing

Mistral 7B

LLaMA

DeepSeek

Bloom





Decision-making framework:

- Define your primary task
- Assess your constraints
- Confirm legal compliance



Thank you for your time!

@bitheap_tech | www.bitheap.ch

